

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт фундаментальной биологии и биотехнологии
Кафедра биофизики

УТВЕРЖДАЮ:
заведующий кафедрой

_____ В. А. Кратасюк
“ ____ ” _____ 2020 г.

БАКАЛАВРСКАЯ РАБОТА

06.03.01 Биология

ИЗУЧЕНИЕ ВНУТРЕННЕЙ СТРУКТУРИРОВАННОСТИ ГЕНОМОВ МИТОХОНДРИЙ НЕКОТОРЫХ СИСТЕМАТИЧЕСКИХ ГРУПП

Руководитель: _____ д.ф.-м.н., проф. М. Г. Садовский
дата, подпись уч.степень, должность

Выпускник: _____ Р. Е. Косарев
дата, подпись

Красноярск 2020

РЕФЕРАТ

Выпускная квалификационная работа состоит из 37 страниц текстового документа, 25 использованных источников, 19 рисунков, 2 таблиц.

СТРУКТУРИРОВАННОСТЬ МИТОХОНДРИАЛЬНЫХ ГЕНОМОВ, ТРЁХЛУЧЕВАЯ СТРУКТУРА, СТРУКТУРИРОВАННОСТЬ РАЗЛИЧНЫХ СИСТЕМАТИЧЕСКИХ ГРУПП, МИТОХОНДРИАЛЬНЫЙ ГЕНОМ, МИТОХОНДРИИ

Цель работы — обнаружение и изучение внутренней структурированности митохондриальных геномов разных систематических групп и описание характерных структур и их различий между исследуемыми группами.

Предмет исследования — нуклеотидные последовательности митохондриальных геномов исследуемых организмов, взятые из открытой базы данных EMBL-банка.

Объект исследования — внутренняя структурированность митохондриальных геномов и их взаимосвязь с соответствующими систематическими группами живых организмов.

Актуальность исследования обусловлена тем, что выявление взаимосвязи структурных компонентов и соответствующих им функций — классическая проблема молекулярной и системной биологии, решение которой позволяет сделать эволюционную картину тех или иных биологических систем точнее.

Отобран и исследован 461 митохондриальный геном живых организмов различных систематических групп. Получены структуры митохондриальных геномов: трехлучевая (т/х), т/х с хвостом, т/х с нимбом, т/х с ядром, клубок и аморфная. Митохондриальные геномы были исследованы по относительным фазам и по GC-составу. Найдены некоторые взаимосвязи между систематическими категориями, GC-составом и окраской по GC-составу.

Содержание

ВВЕДЕНИЕ	4
1 Обзор литературы	7
1.1 Митохондрии	7
1.2 Геномы митохондрий	9
1.3 Изучение структурированности в других исследованиях . . .	10
2 Материалы и методы	12
2.1 Отбор генетического материала митохондрий	12
2.2 Частотные словари	13
2.3 Программное обеспечение VidaExpert	15
2.4 Окрашивание точек по GC-составу	16
3 Результаты	18
3.1 Распределение систематических групп по GC-составу	18
3.2 Структурированность кодирующих и некодирующих последовательностей митохондриальных геномов	21
3.3 Взаимосвязь GC-состава и структурированности митохондриальных геномов	27
3.4 Окрашивание полученных структур по GC-составу	28
Заключение	33
Список сокращений	34
Список использованных источников	35

ВВЕДЕНИЕ

В настоящее время изучение структурированности нуклеотидных последовательностей является одной из важнейших задач биологии. Подобные исследования рассматриваются в структурно-функциональном и эволюционном аспектах. Выявление взаимосвязи структурных компонентов и соответствующих им функций является классической проблемой молекулярной и системной биологии. Даже несмотря на огромное количество публикаций и исследований в этом направлении, которые постоянно дополняются, она всё ещё далека от завершения. Помимо этого, исследователи обнаруживают структурные элементы, которые ранее было трудно выявить, или же такие виды и формы взаимодействий и взаимоотношений между структурными элементами биологических макромолекул, которые на другом этапе развития были бы недоступны для исследования. В данном контексте стремительное увеличение количества и качества технического оснащения и инструментов исследований

Подобные исследования чрезвычайно важны с точки зрения понимания эволюционных процессов. Изучение структурных особенностей биологических макромолекул, как у разных организмов, так и у систематических групп в целом, дает возможность сделать эволюционную картину тех или иных биологических систем точнее. Такими системами могут быть как конкретные виды, так и целые экосистемы, и глобальные сообщества.

Однако, трудности в подобных исследованиях вызывают выбор и качество рассматриваемого биоматериала. Суть проблемы не в ошибках аннотирования и/ли секвенирования генетических последовательностей, которые трудно избежать во многих случаях, а в большой сложности геномов и отдельных хромосом. При рассмотрении данных объектов необходимо анализировать следующие характеристики: структурированность, функции и филогению. Эти характеристики, как правило, глубоко взаимосвязаны и оказывают друг на друга сильное влияние. Сложность в том, что это влияние редко возможно выделить в качестве отдельного и независимого фактора.

С точки зрения этой проблемы прокариотические организмы являются более удобными объектами для исследования, чем эукариотические, так как геном бактерий значительно короче генома эукариот и он почти всегда представлен одной хромосомой, за исключением некоторых организмов, например, *Rhodobacter sphaeroides* [1]. Как итог, самым удобным объектом исследования является геном органелл, таких как митохондрии и хлоропласты. В нашей работе — это митохондрии. Такой выбор объясняется тем, что для этих органелл полностью исключается фактор влияния различий в кодируемых функциях, так как в пределах одной группы органелл, будь то митохондрий или хлоропласт, функциональные различия минимальны.

Объектом исследования в настоящей работе являются типы и виды внутренней структурированности митохондриальных геномов различных систематических групп царства животных:

- подтип позвоночные: классы — круглоротые, хрящевые рыбы, костные рыбы, амфибии, птицы, млекопитающие;
- типы: губки, моллюски, плоские черви, круглые черви, кольчатые черви, членистоногие;
- последние делятся на классы: насекомые, паукообразные, ракообразные.

Цель настоящей работы — выявление и исследование внутренней структурированности митохондриальных геномов и описание характерных структур и их различия между разными систематическими группами. Для достижения указанной цели были поставлены следующие **задачи**:

- написать литературный обзор по митохондриям и изучению структурированности;
- исследовать закономерности между GC-составом и изучаемыми систематическими группами;
- выявить различные внутренние структуры митохондриальных геномов, характерные для исследуемых систематических групп;
- классифицировать полученные структуры;

- изучить взаимосвязь полученных структур и их отдельных компонентов с окраской по **ГС**-составу.

1 Обзор литературы

1.1 Митохондрии

Митохондрии — клеточные органеллы палочкообразной формы, диаметр которых приблизительно равен 1 мкм, а длина около 7 мкм. Данные органоиды находятся исключительно в клетках эукариот. Митохондрии — двухмембранные органеллы. Складки внутренней мембраны называются кристами, тогда как внутри находится митохондриальный матрикс, содержащий рибосомы и митохондриальную ДНК. Эта нуклеиновая кислота представляет собой кольцевую двухцепочечную молекулу, функция которой — сохранение и реализация информации о некоторых митохондриальных белках.

За синтез АТФ отвечает F₁ комплекс, который является ферментом, располагающимся во внутренней мембране. Между наружной и внутренней мембранами располагаются ферменты, субстраты, и некоторые метаболиты. В различных клетках находится до нескольких тысяч митохондрий. Главной функцией этих органелл является окисление органических соединений с последующим синтезом АТФ, для чего используется энергия окисляющихся веществ. Главные компартменты задействованные в этом — кристы митохондрий. Они участвуют в системе переноса электронов.

Молекула АТФ — универсальный клеточный источник энергии, несмотря на то, что в некоторых реакциях возможно использование других нуклеозидтрифосфатов, таких как ГТФ. АТФ производится в цитоплазме. Для этого происходит ряд катаболических процессов, например, гликолиз, а энергия, запасенная в трифосфатной группе, далее идёт в расход в ходе реакций биосинтеза.

Сходство генетической системы митохондрий и хлоропластов указывает на возможное происхождение этой органеллы в ходе эволюции из бактерий, которые были эндоцитированы более одного миллиарда лет назад. Эта гипотеза называется эндосимбиотической. Ее смысл заключается в том в том, что эукариотические клетки вначале были анаэробными организмами, в них отсутствовали митохондрии, после чего они возможно смогли

вступить в стабильные эндосимбиотические отношения с бактерией. В ходе данного процесса митохондрии приспособились к системе окислительного фосфорилирования. Судя по анализу сравнения последовательностей можно предположить, что предковой формой митохондрии может быть определенный тип пурпурных фотосинтетических бактерий, которые до этого потеряли способность к фотосинтезу, в следствие чего у них осталась только дыхательная цепь [2]. Наиболее вероятным предком современных митохондрий считают альфа-протеобактерии с окислительной системой в дыхательной цепи [3].

С течением времени они потеряли большое количество изначального ДНК, так как их гены постепенно были перемещены и включены в ядерный геном хозяев, обитателями которых они являлись. Однако они по-прежнему способны поддерживать собственную ДНК. ДНК современных митохондрий кодирует только часть белков, которые необходимы для дыхания, фотосинтеза и других функций органелл. Остальные белки, нужные для полноценного функционирования органеллы, кодируются в ядре и в ходе клеточных процессов транспортируются в митохондрии. Они отвечают за механизмы, необходимые для репликации ДНК, транскрипции генов, кодирования некоторых белков и тРНК, для дальнейшей реализации этой кодируемой информации в ходе трансляции.

Митохондриальные геномы растений варьируют от 200 до 2000 п.н. (пар нуклеотидов) в длину, тогда как почти все митохондриальные геномы животных составляют около 16,5 п.н. Однако несмотря на такую большую разницу в размерах, растения не кодируют значительно большее число генов в своих митохондриальных геномах, чем животные. Большая часть митохондриальной ДНК растительных организмов состоит из крупных интронов, повторов и некодирующих областей. Также в растительных митохондриальных геномах, в связи с наличием в растениях хлоропластов, содержатся значительные количества относительно коротких хлоропластных и ядерных геномных последовательностей.

У растений фрагментированная природа митохондриальных геномов

приводит к частым рекомбинациям ДНК. Одна митохондрия, как правило, обладает множеством копий собственного генома. Количество копий на органеллу варьируется от 50 до 500 и зависит от возраста растения и от типа ткани. Исследования на кукурузе показали, что с увеличением возраста растений число копий мтДНК постепенно уменьшается.

Митохондрии способны сливаться вместе, образуя при этом крупные трубчатые сети, из-за чего у них появляется способность легко переносить генетический материал из одной митохондрии в другую. Из-за этого некоторые митохондрии могут находиться в одной и той же клетке с разным количеством копий генома.

1.2 Геномы митохондрий

Как правило митохондриальная ДНК — кольцевая двухцепочечная молекула. Иногда обнаруживаются линейные ДНК, такие как у рода паразитических одноклеточных организмов *Plasmodium*. В матриксе митохондрий располагается значительное количество копии ДНК. Они, как правило, бывают распределены по нескольким кластерам — нуклеоидам. Обычно нуклеоиды прикреплены к внутренней мембране.

Пространственная структура нуклеоидной ДНК больше походит на структуру бактериальной ДНК, чем эукариотический хроматин. Один из признаков их близости — отсутствие в органеллах гистонов. Разница в размерах молекул митохондриальной ДНК составляет от менее 600 пар до 300 тысяч п. н. (в малярийном паразите человека (*Plasmodium falciparum*) и некоторых наземных растениях соответственно) [4, 5]. Общим признаком митохондриальных и ядерных геномов эукариот является большое количество разбросанных по их геному некодирующих последовательностей.

Экзоны — участки генома, которые составляют кодирующую часть гена. Интроны — некодирующие участки, которые располагаются между экзонами и не содержат ген. Помимо этого, между генами может находиться так называемая «мусорная» ДНК, которая состоит из некодирующих участков генома. Такое название объясняется отсутствием доказательств

явных полезных функций для клетки. Несмотря на это, небольшой частью этой ДНК реализуется экспрессия некоторых генов.

Реализация генетического потенциала происходит в ходе процесса, который называется сплайсингом. Он происходит следующим образом:

- С каждого гена синтезируется пре-РНК, которая состоит из интронов и экзонов;
- Интроны удаляются, а экзоны сшиваются в одну цепь зрелой иРНК.

Митохондриальные геномы грибов и растений иногда содержат интроны, что является редкостью, поскольку митохондрии, как предполагается, произошли от симбиоза с бактериями, у которых обычно в генах интроны отсутствуют. Так же у дрожжей иногда случается так, что в одном и том же митохондриальном гене может содержаться интрон в одном штамме, в то время как в другом штамме он будет отсутствовать. Возможно, такие интроны могут покидать геном и возвращаться обратно, как мобильные генетические элементы.

1.3 Изучение структурированности в других исследованиях

О влиянии структуры митохондриальных геномов разных организмов на их филогению и эволюцию написано множество различных работ [6–19]. Эти работы объединяет общий подход в том, что они рассматривают структуру митохондриальных геномов относительно функций соответствующих генов.

Важной особенностью нашей работы является исключение функциональной зависимости из исследования геномов митохондрий. Мы рассматриваем структуру митохондриальных геномов как кластеризацию точек данных, которые соответствуют частотным словарям в пространстве триплетов трех главных компонент. Структурированность — это кластеризация частотных словарей триплетов отдельных генетических участков, которая определяется регулярным порядком и не зависит от функциональной роли того или иного участка. Сам подход в исследовании структурирован-

ности митохондриальных геномов будет сильно схож с соответствующим подходом в исследовании структурированности геномов хлоропластов в работе [20].

В этой работе исследовалась внутренняя структурированность геномов хлоропластов и цианобактерий, с целью обнаружить схожесть или различия со структурированностью бактериальных геномов, которые были исследованы в ряде работ [21–23]. Результаты работ показали, что структурированности значительно отличаются от ранее обнаруженной для геномов бактерий.

2 Материалы и методы

2.1 Отбор генетического материала митохондрий

Данные для исследования брались из открытой базы данных EMBL-банка содержащей все проверенные и аннотированные на сегодняшний день последовательности ДНК митохондрий [24]. Было отобрано 461 митохондрия живых организмов, в среднем по 25 видов из каждой систематической группы. В таблице 1 указано количественное соотношение исследуемой выборки организмов по систематическим группам.

Таблица 1 – Количественное соотношение исследуемой выборки организмов по систематическим группам.

систематическая категория	систематическая группа	N, количество
тип	Губки	24
тип	Моллюски	25
класс	Амфибии	25
класс	Круглоротые	14
класс	Хрящевые рыбы	25
класс	Костные рыбы	25
тип	Плоские черви	25
тип	Круглые черви	25
тип	Кольчатые черви	25
класс	Ракообразные	25
класс	Паукообразные	25
класс	Насекомые	25
класс	Рептилии	24
класс	Птицы	25
класс	Млекопитающие	25
царство	Растения	35
царство	Грибы	25
отдел	Мхи	25
отдел	Лишайники	13

2.2 Частотные словари

В ходе данного исследования для осуществления метода рассматривалась генетическая последовательность длины L . Такая последовательность включает в себя следующие символы алфавита $\aleph = \{A, C, G, T\}$. Для нее создавался частотный словарь толщины 3. Частотный словарь W_3 толщины 3 символьной последовательности, соответствующей ДНК — это список всех троек $\nu_1\nu_2\nu_3$ расположенных друг за другом нуклеотидов с обозначением частоты каждой тройки в рассматриваемом участке. Так как практически любой геном кодируется четырьмя нуклеотидами, соответственно частотный словарь включает в себя 64 триплета. Стоит упомянуть, что данное определение частотного словаря триплетов — частный случай, так как окно считывания способно перемещаться на любое заданное число нуклеотидов во время подсчета числа триплетов. Тем самым возможно использовать любой другой частотный словарь, если он подходит под объекты исследования и дает исследователю максимум информации об исследуемой структуре. Принято использовать частотный словарь вида W_3^k , где $k = 1$.

В нашей работе используется частотный словарь следующего вида — W_3^3 . В нашей работе триплеты подсчитывались так, что они, не пересекаясь, полностью покрывали генетическую последовательность. Частота, обозначаемая f_ω — это отношение числа копий n_ω выбранного триплета к общему числу всех триплетов N в исследуемом участке, где N — сумма всех n_ω :

$$f_\omega = \frac{n_\omega}{N}. \quad (1)$$

Каждый исследуемый частотный словарь W_3^3 переводит геном в 64-мерное пространство. Сумма всех частот в словаре равняется единице, что приводит к линейной связи, которая может дать ложный сигнал в результате статистической обработки (определении главных компонент, корреляционном анализе и т. п.). Поэтому один из триплетов желательно исключать.

Мы исключаем триплет с наименьшим стандартным отклонением по исследуемому набору частотных словарей. Триплет с минимальным стан-

дартным отклонением дает наименьший вклад в различимость объектов. После этого шага рассматриваемое пространство точек из 64-мерного переходит в 63-мерное.

Для выявления структуры в генетической последовательности проводилась предварительная обработка, которая ставила в соответствие данной последовательности множество точек в 63-мерном пространстве триплетов. Делалось это следующим образом: последовательность сканировалась рамкой считывания длины Δ с шагом t . Для каждого положения i рамки определялся участок генетической последовательности, совпадающий с рамкой считывания, для которого вычислялся частотный словарь $W_3^{(i)}$ соответствующий i -ой точке в 63-мерном пространстве. Кроме того, с каждой точкой в 63-мерном пространстве связывались следующие параметры: номер центрального символа рассматриваемого участка и относительная фаза. Номер центрального символа участка совпадает с номером этого символа в последовательности.

Относительная фаза определяется в зависимости от того, попал рассматриваемый участок в кодирующую или не кодирующую область последовательности. Участок относится к кодирующим, если он целиком попал в кодирующую область последовательности. Если участок относится к не кодирующим, то соответствующая ему точка помечается символом J . Для кодирующего участка возможны шесть вариантов маркировки: $B_0, B_1, B_2, F_0, F_1, F_2$. Если кодирующий участок в последовательности аннотирован как считывающийся в прямом направлении, то для него вычислялся остаток от деления на 3 разности номеров центрального символа участка и первого символа кодирующей области, к которой он относится. В соответствии с величиной остатка от деления точка помечалась символом F_0, F_1 или F_2 . Если участок аннотирован как считывающийся в обратном направлении, то вычислялся остаток от деления на 3 разности номеров последнего символа кодирующей области, к которой относится участок, и центрального символа участка. В зависимости от значения остатка от деления точка помечалась символами B_0, B_1 или B_2 .

2.3 Программное обеспечение VidaExpert

По полученному множеству точек в программе VidaExpert¹ строился вид данных в пространстве первых трех главных компонент, вычисленных для данного 63-мерного пространства. Рассматривались две проекции на плоскость пространства главных компонент: 1-ой и 2-ой компоненты и 2-ой и 3-ей компоненты (в исключительных случаях рассматривалась проекция на плоскость пространства 1-ой и 3-ей компонент для наглядности). Точки данных были раскрашены в соответствии с их относительной фазой. Точки, соответствующие некодирующим участкам, закрашены коричневым, точки, соответствующие относительным фазам F_0 и B_0 , изображены малиновыми и фиолетовыми квадратами, точки, соответствующие относительным фазам F_1 и B_1 , изображены зелеными и голубыми квадратами, а точки, соответствующие относительным фазам F_2 и B_2 , изображены оранжевыми и желтыми квадратами. Кроме того, для каждого генома вычислялся GC-состав, т.е. отношение суммарного числа нуклеотидов С и G к общему числу нуклеотидов в геноме и находилось среднее значение GC-состава по кладе.

Для облегчения восприятия результатов введем точные определения терминов, которые будут использоваться ниже. Выделяемые рамкой считывания длины Δ фрагменты помечались двумя разными метками: ген и джанк. Следует подчеркнуть, что фрагменты выделяются только из одной (прямой) цепи ДНК и каждый такой фрагмент может содержать: 1) кодирующую область из прямой цепочки с тремя возможными сдвигами относительно первой позиции фрагмента; 2) кодирующую область из обратной цепочки с тремя возможными рамками считывания (определяемые по “зеркальному отражению”); 3) некодирующие области; и, наконец, 4) сочетание некодирующих и кодирующих областей в разных пропорциях. Эти последние фрагменты вносят шум в распределение, но их относительная доля не высока. Понятно, что выбор порога доли кодирующей области в выделяемом фрагменте будет оказывать некоторое влияние на результаты распределения; в рамках настоящей работы использовался весьма жесткий

¹<http://bioinfo-out.curie.fr/projects/vidaexpert/>

критерий: в категорию “ген” попадали лишь те фрагменты, которые целиком лежали в кодирующей области. Это обусловлено тем, что в противном случае выбор доли кодирующих и не кодирующих участков во фрагментах, которые будут считаться “геном” очень субъективен. Влияние величины доли не кодирующих областей в выделяемых фрагментах нами не исследовалось.

Фрагмент с номером i получал метку ген, если он попадал в ту часть генома, которая в аннотации, содержащейся в голове файла EMBL-банка, была помечена как CDS, либо как exon, либо как rRNA, если длина такового превышала длину фрагмента. При этом фрагмент получал такую метку лишь при условии полного попадания в участок, аннотированный как описано выше. Фрагмент с номером i получал метку джанк, если он не получал метку ген, при условиях, описанных выше.

Как показано в работе [20], у хлоропластов наземных растений преобладающей была обнаружена трёхлучевая структура с хвостом. В нашем исследовании объектом исследования являются митохондрии животных и растений, чей геном значительно короче генома хлоропластов. Данное обстоятельство может быть причиной обнаруженных различий.

2.4 Окрашивание точек по GC-составу

Для работы с участками генетических последовательностей в программе VidaExpert каждая точка, отвечающая за определенный участок, окрашивалась в соответствии с GC-составом этого конкретного участка, где GC-состав — это процентное содержание нуклеотидов гуанина и цитозина среди всех нуклеотидов в соответствующем участке.

Окраска производится для каждого конкретного вида независимо. Окраску точек по GC-составу можно разделить на 6 этапов, начиная с построения структуры:

1. Открыть программу VidaExpert → вкладка Project → Load table → выбрать соответствующий митохондриальному геному изучаемого вида dat-файл → в колонке с триплетами клик правой кнопкой мыши Select

- all → повторный клик правой кнопкой мыши Mark → убрать галочки в строке «GC» и строке триплета с минимальным стандартным отклонением;
2. Вкладка Data → Create dataset → в колонке от появившегося значка «глаза» выбрать параметр Soft map 16*16 → нажать на значок «глаза»;
 3. В предпоследней колонке справа выбрать соответствующий Datasets «XX», полученный на 2 этапе, → выбрать подходящую карту «Map [тип/размер]» (чаще 001) → в крайней колонке справа во Views выбрать параметр On principal components;
 4. Иконка Statistics в верхней панели → вкладка Histograms → в колонке слева выбрать GC → в поле Bars выбрать значение «6» → оценить и запомнить полученные значения по получившимся рядам данных и расстояния между ними, которые необходимо далее использовать;
 5. Иконка Object selection → поле Field выбрать GC → значение меньше «<» → выбрать максимальное численное значение GC-состава в следующем поле → для максимума используется Shape (фигура) — квадрат, Fill color (цвет) красный, Size для всех случаев «5»;
 6. Таким же образом выбрать меньшее численное значение GC, которое доходит до границы следующего ряда данных по диаграмме из этапа 4. Выбрать форму и окрасить по относительному содержанию GC-состава в следующем порядке:
 - Красный квадрат (~80-100%);
 - Оранжевый треугольник (~65-80%);
 - Желтый треугольник (~50-65%);
 - Зелёный треугольник (~35-50%);
 - Голубой треугольник (~20-35%);
 - Синий квадрат (~0-20%).

Как видно из используемых форм и цветов, для крайних значений используются квадраты, для всего остального треугольники, во избежание загруженности конечного вида структуры.

3 Результаты

3.1 Распределение систематических групп по GC-составу

Мы начнем представление результатов с описания базы исследованных геномов. В каждом геноме выделялись два вида участков: ген и джанк (их детальное описание см. выше). Кроме того, был определен GC-состав для всех выбранных систематических групп.

Так, на рисунке 1 представлено распределение систематических групп по GC-составу, в направлении эволюционного развития от простых жизненных форм, представленных этими группами, до сложных, таких, как млекопитающие и птицы. В построении участвовал 461 вид из 19 систематических групп.

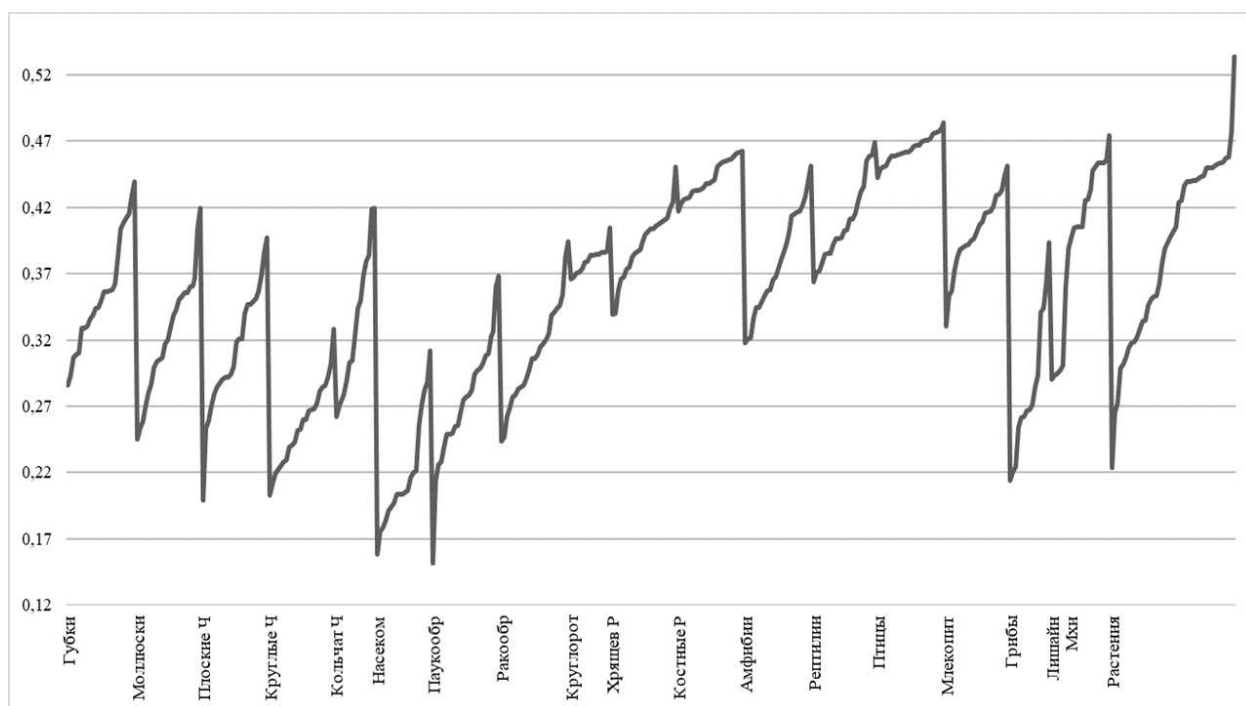


Рисунок 1 – GC-состав исследуемых систематических групп. Каждая систематическая группа ранжирована по возрастанию содержания GC

Уже на рис. 1 можно определить некоторые закономерности, характерные для различных систематических групп. Так же, по сравнению с результатами работы [20] заметна бóльшая нестабильность изучаемых геномов митохондрий по GC-составу, как среди животных, так и среди растений, нежели у 185 геномов хлоропластов, представленных в его работе.

Это, так же, может объясняться разницей в размерах геномов хлоропласт и митохондрий.

Для более удобного восприятия данных по **ГС**-составу мы представляем также несколько диаграмм в формате «ящик с усами» систематических групп с близким содержанием **ГС**-состава рис. 2-6.

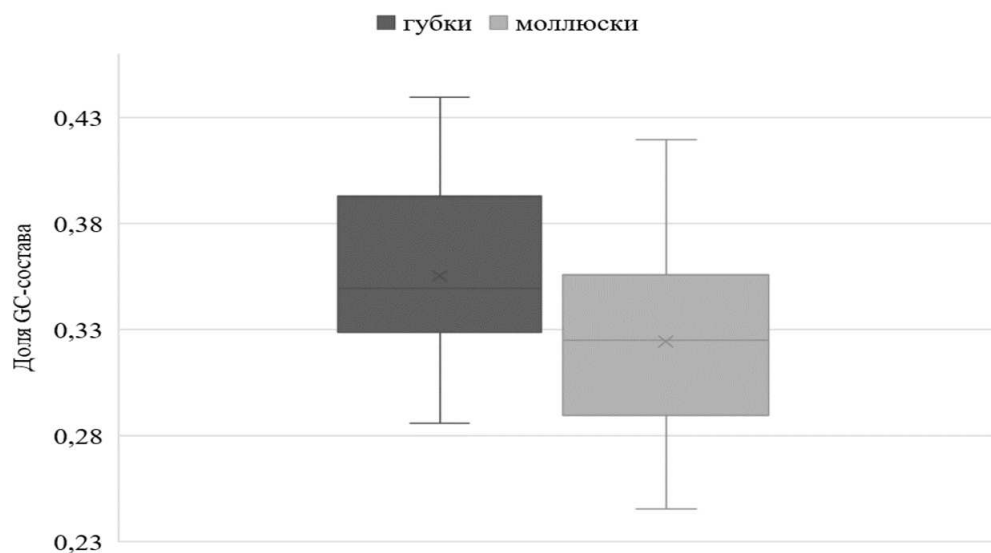


Рисунок 2 – **ГС**-состав водных беспозвоночных

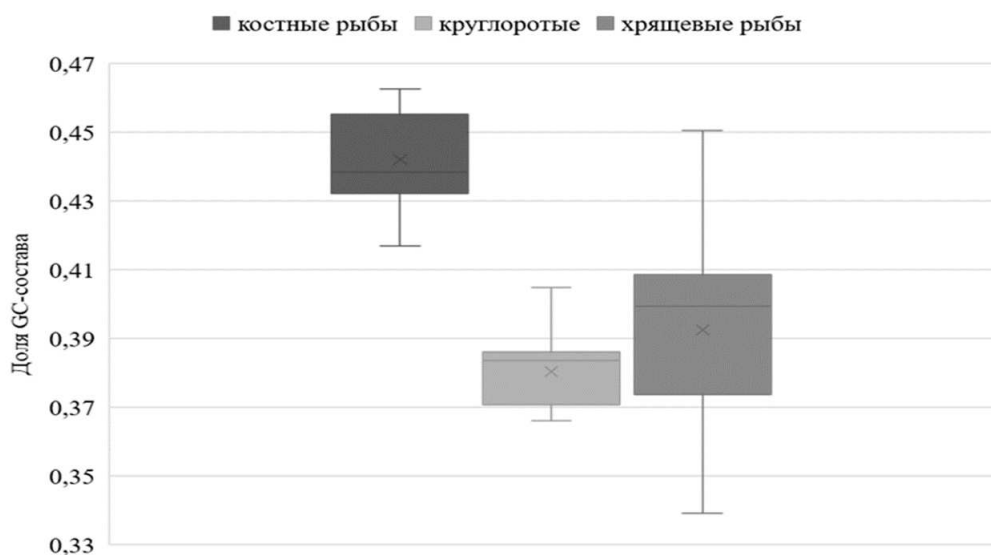


Рисунок 3 – **ГС**-состав водных позвоночных

На данном этом этапе можно объединить некоторые систематические категории по родству на основе **ГС**-состава как на рис. 2-6.

Так выделяются следующие группы:

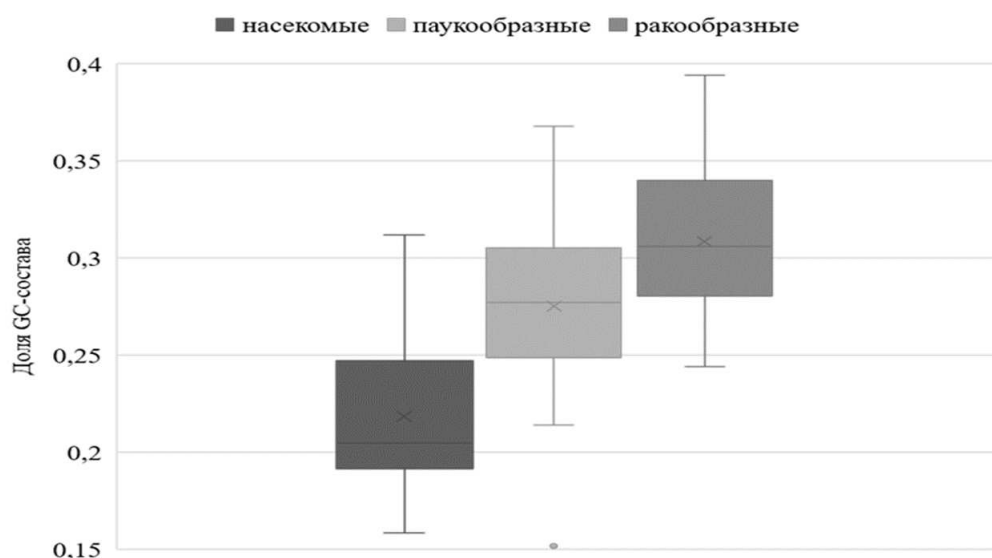


Рисунок 4 – GC-состав наземных беспозвоночных (членистоногих)

- Водные беспозвоночные (губки и моллюски);
- Водные позвоночные (круглоротые, хрящевые и костные рыбы);
- Наземные беспозвоночные, черви (плоские, круглые и кольчатые);
- Наземные беспозвоночные, членистоногие (насекомые, ракообразные, паукообразные);
- Наземные позвоночные (амфибии, рептилии, птицы, млекопитающие) и др.

Полученные данные можно объяснить филогенией этих систематических групп и их эволюционной близостью.

Самое низкое среднее значение GC-состава наблюдается у митохондриальных геномов круглых червей (как представителей наземных беспозвоночных, червей в целом): *Trichuris suis*, *Trichuris ovis*, *Angiostrongylus cantonensis*, *Angiostrongylus costaricensis*, *Cooperia oncophora*, *Metastrongylus pudendotectus*, *Metastrongylus salmi*, *Mecistocirrus digitatus*, *Setaria digitata*, *Oesophagostomum columbianum*, *Gongylonema pulchrum*, *Caenorhabditis tropicalis*, *Strongylus equinus*, *Pseudoterranova azarasi*, *Onchocerca volvulus*, *Haemonchus placei*, *Triodontophorus serratus*, *Triodontophorus nipponicus*, *Cylicocyclus nassatus*, *Setaria digitate*, *Trichinella spiralis*, *Dracunculus medinensis*, *Ascaris lumbricoides*, *Passalurus ambiguus*, *Ortleppascaris sinensis*.

В то время, как самое высокое среднее значение GC-состава наблю-

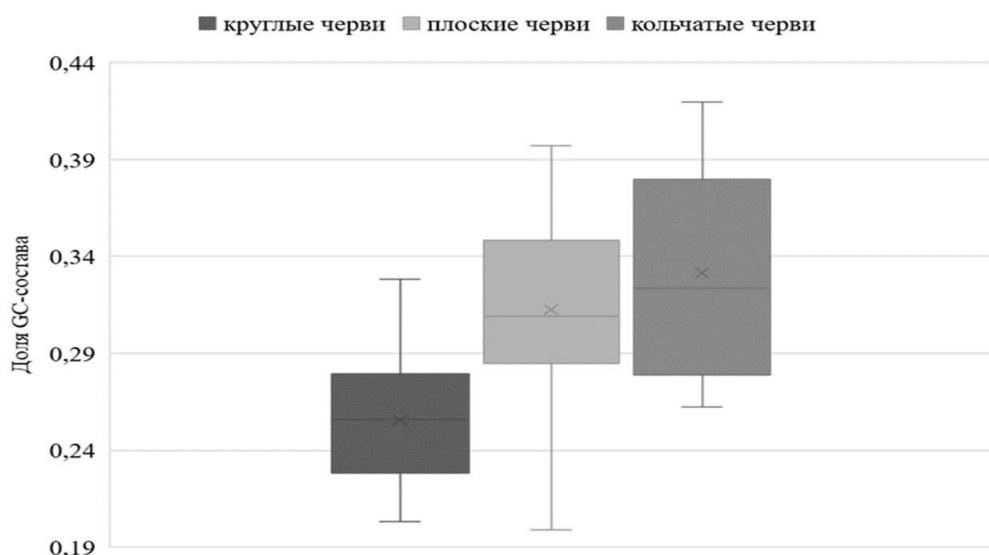


Рисунок 5 – GC-состав наземных беспозвоночных (червей)

дается у митохондриальных геномов птиц (как у представителя наземных позвоночных): *Luscinia calliope*, *Cyanoptila cyanomelana*, *Acridotheres cristatellus*, *Gracula religiosa*, *Petroica boodang*, *Phoenicurus auroreus*, *Luscinia cyanura*, *Turdus eunomus*, *Periparus ater*, *Cygnus cygnus*, *Napothera epilepidota*, *Streptopelia decaocto*, *Streptopelia orientalis*, *Corvus brachyrhynchos*, *Gallus gallus*, *Columba livia*, *Falco peregrinus*, *Ara severus*, *Caprimulgus indicus*, *Circus melanoleucos*, *Chaetura pelagica*, *Dendrocopos leucotos*, *Aquila heliaca*, *Falco tinnunculus*, *Ardea intermedia*.

3.2 Структурированность кодирующих и некодирующих последовательностей митохондриальных геномов

На данном этапе мы переходим к описанию структурированности, выделенной в геномах митохондрий различных систематических групп. Часть полученных данных была представлена ранее в работе [25].

Все дальнейшие структуры рассматриваются в двух проекциях, среди которых вид на плоскость 1 и 2 главных компонент (левое изображение) и вид на плоскость 2 и 3 главных компонент (правое изображение), подробнее можно рассмотреть на рис.7

У 19 рассмотренных систематических групп живых организмов выявляются 6 основных типов структур, см. рис 8 – 13.

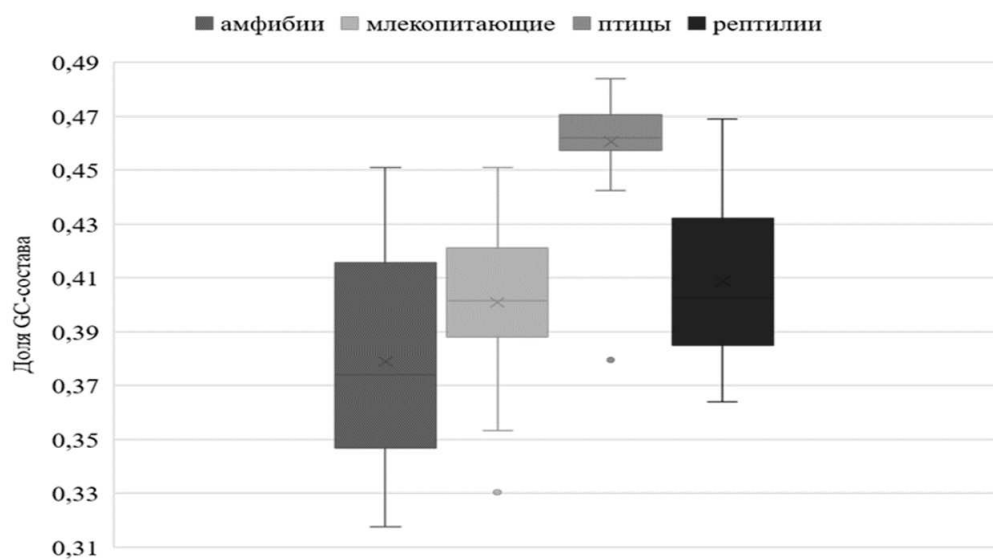


Рисунок 6 – GS-состав наземных позвоночных

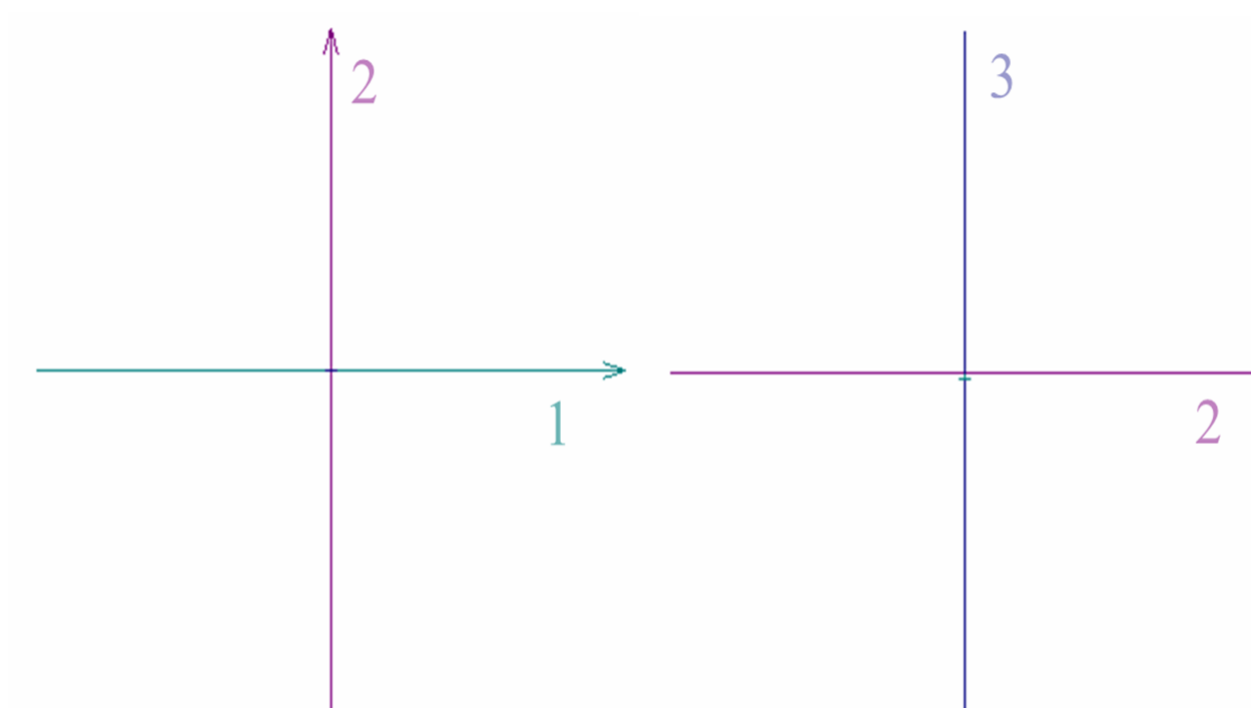


Рисунок 7 – Рассматриваемые плоскости пространств трех главных компонент. Слева плоскость 1 и 2 главных компонент. Справа плоскость 2 и 3 главных компонент

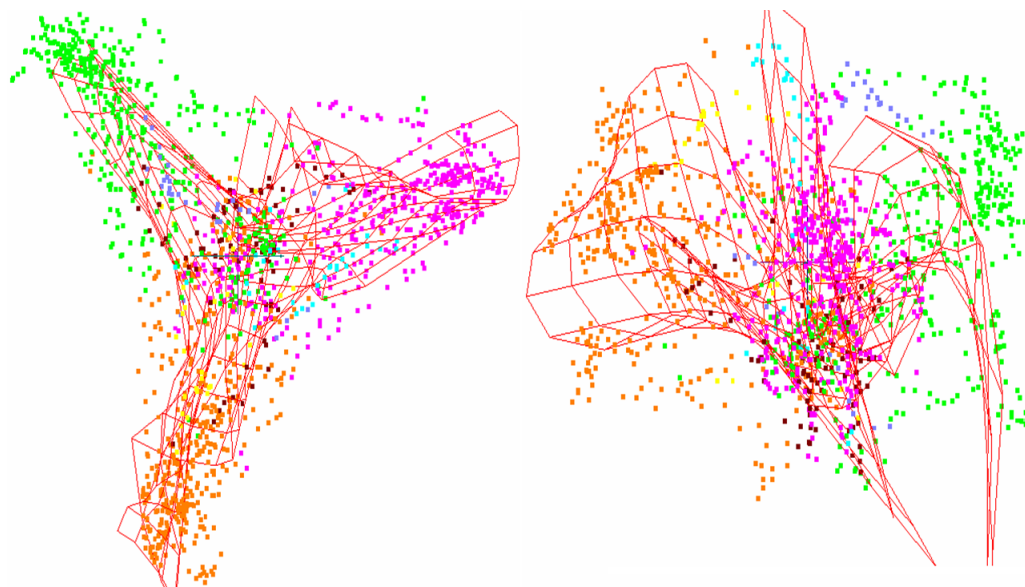


Рисунок 8 – Трехлучевая структура на примере *Coelodonta antiquitatis* (млекопитающие).

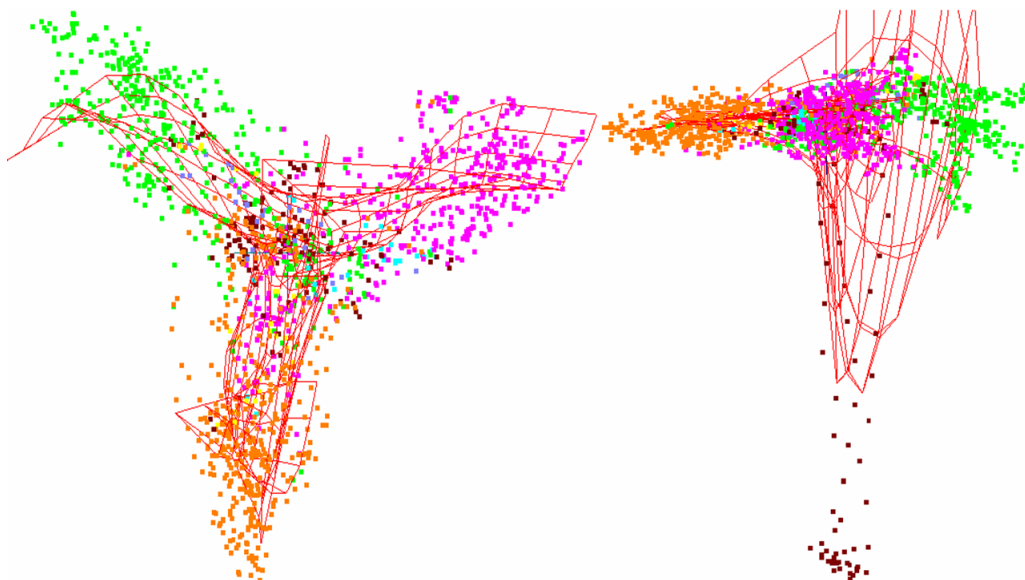


Рисунок 9 – Трехлучевая структура с хвостом на примере *Cuon alpinus lepturus* (млекопитающие).

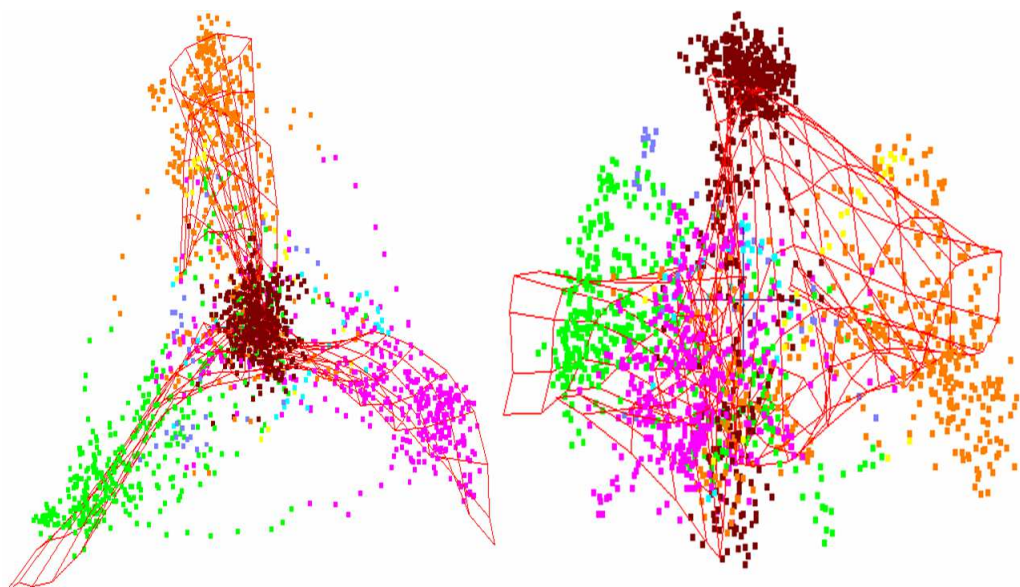


Рисунок 10 – Трехлучевая структура с на примере *Chimaera fulva* (хрящевые рыбы).



Рисунок 11 – Трехлучевая структура с ядром на примере *Ganoderma lucidum* (высшие грибы).

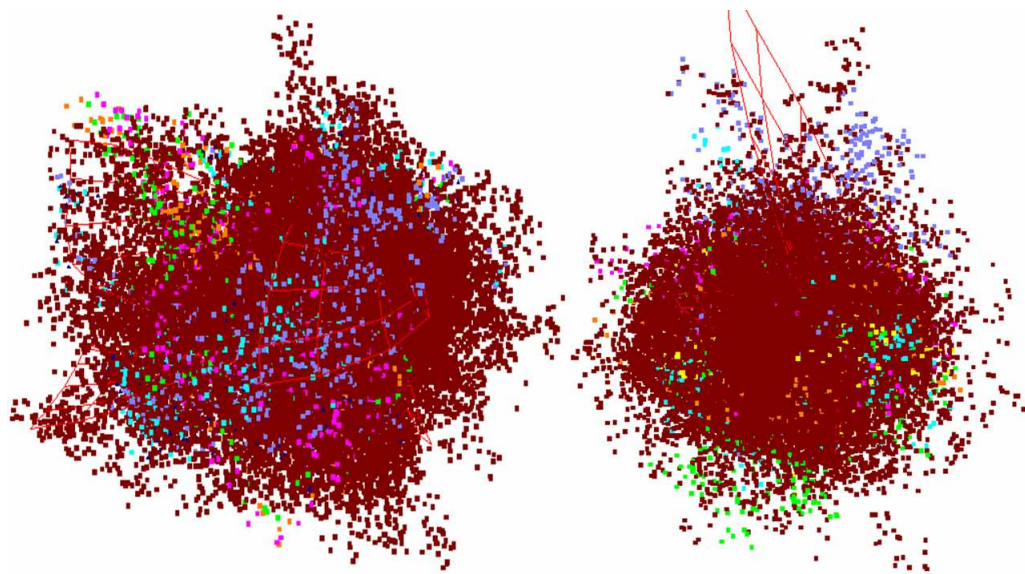


Рисунок 12 – Структура клубок на примере *Oryza rufipogon* (растения).

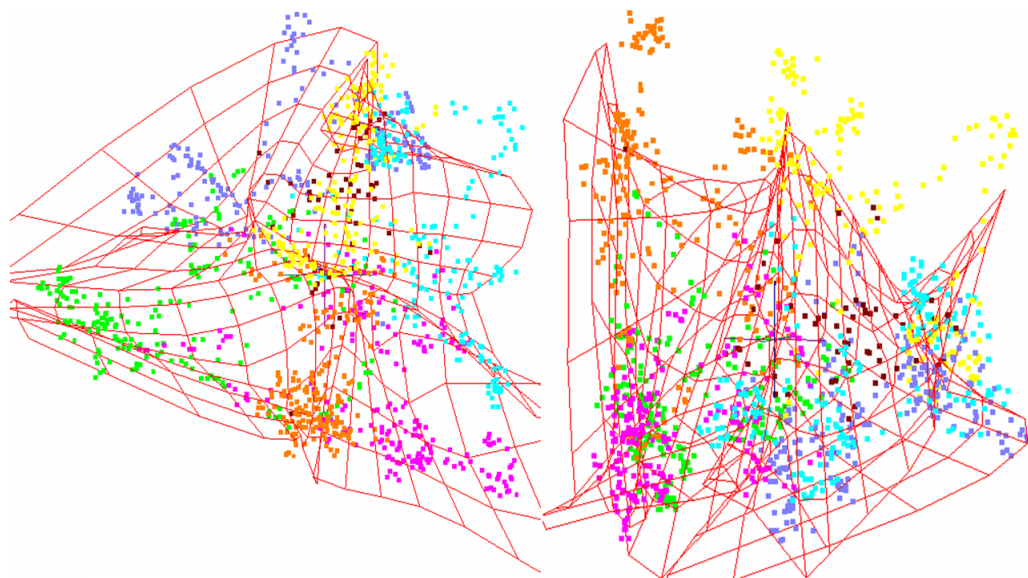


Рисунок 13 – Аморфная структура на примере *Heterometrus longimanus* (паукообразные).

Количественное и процентное соотношение структур описано в таблице 2.

Таблица 2 – Распределение выявленных структур по систематическим группам в процентном соотношении из всей выборки организмов по конкретной систематической группе. Т/х — трехлучевая; Амф — аморфная

Сист. группа	Тип структуры					
	Т/х (Т/х)	Т/х с хвостом	Т/х с нимбом	Т/х с ядром	Клубок	Амф
Губки	23 (92%)	1 (4%)	0	0	0	1 (4%)
Моллюски	19 (76%)	1 (4%)	2 (8%)	0	0	3 (12%)
Амфибии	24 (96%)	1 (4%)	0	0	0	0
Круглоротые	12 (86%)	2 (14%)	0	0	0	0
Хрящевые рыбы	15 (60%)	2 (8%)	8 (32%)	0	0	0
Костные рыбы	21 (84%)	0	4 (16%)	0	0	0
Плоские черви	21 (84%)	3 (12%)	0	0	0	1 (4%)
Круглые черви	18 (72%)	5 (20%)	0	0	0	2 (8%)
Кольчатые черви	23 (92%)	1 (4%)	1 (4%)	0	0	0
Ракообр.	19 (76%)	0	6 (24%)	0	0	0
Паукообр.	18 (72%)	0	2 (8%)	0	0	5 (20%)
Насекомые	19 (76%)	0	6 (24%)	0	0	0
Рептилии	13 (52%)	12 (48%)	0	0	0	0
Птицы	16 (64%)	9 (36%)	0	0	0	0
Млекопит.	19 (76%)	5 (20%)	0	0	0	1 (4%)
Растения	12 (34,3%)	3 (8,6%)	0	0	17 (48,5%)	3 (8,6%)
Грибы	10 (40%)	0	0	14 (56%)	0	1 (4%)
Мхи	0	0	0	11 (44%)	14 (56%)	0
Лишайники	13 (100%)	0	0	0	0	0

В таблице 2 показано, что для большинства систематических групп типичной является трехлучевая структура. Если не рассматривать отдельно подтипы трехлучевой структуры, то кроме как у некоторых растений и печеночных мхов она является основной для всех остальных систематических категорий.

Трехлучевая структурированность другого вида обнаруживается в [20] с геномами хлоропластов. Больше всего на трехлучевые структуры хлоропластов похожи структуры трехлучевая с ядром (в большом коли-

честве обнаружена у высших грибов и мхов) и трехлучевая с хвостом (в основном, встречается у птиц и рептилий).

Если сравнивать эти структуры с обнаруженными у геномов хлоропластов, то для хлоропластов характерно практически равное количество точек, относящихся к прямым и обратным фазам, тогда как у митохондрий сильно преобладают точки, относящиеся к прямой фазе. Иногда точки, относящиеся к обратной фазе, вовсе отсутствуют (у плоских червей, высших грибов и лишайников). Также «хвост» у хлоропластов состоит из точек, относящихся к кодирующим участкам, в то время как у митохондрий хвост состоит исключительно из джанка. Поэтому говорить об идентичности структур геномов митохондрий и хлоропластов не представляется возможным.

Принципиально новой структурой, выделенной в филогенетически родственных систематических группах, является — трехлучевая с нимбом. У этой структуры в проекции 1-ой и 2-ой главных компонент можно обнаружить сформированный «нимб» вокруг структуры из кодирующих участков. Она представлена у костных и хрящевых рыб, а также у членистоногих (насекомых, ракообразных и немного у паукообразных). Этот вопрос требует дальнейшего изучения в следующих работах. Некоторые структуры получились аморфными, без каких-либо общих признаков и/ли ярко выраженной симметрии. Это можно объяснить небольшими размерами данных геномов, либо ошибками в их аннотации (чаще всего такая структура встречается у паукообразных). Для растений и мхов самой характерной структурой является «клубок» (41% и 56% соответственно). Она выглядит как беспорядочно перемешанное скопление точек.

3.3 Взаимосвязь GC-состава и структурированности митохондриальных геномов

При попытке соотнести GC-состав (см. выше рис 1-6) с полученными структурами (см. выше рис. 8-13) выявляются слабовыраженные закономерности. У птиц и рептилий самая высокая встречаемость структуры

трехлучевой с хвостом, чуть меньше она встречается у млекопитающих. У костных рыб встречается трёхлучевая структура с нимбом. Она также представлена в систематических категориях с наименьшим GC-составом (членистоногие). У мхов часто встречающаяся трехлучевая структура с ядром, которая ещё более характерна для высших грибов, у которых GC-состав ниже среднего.

К сожалению, считать данные закономерности четкой корреляцией между GC-составом и преобладанием какой-либо структурой у геномов митохондрий изучаемых систематических групп не представляется возможным. В таком случае, выборка большего объема, например, не 25 видов на систематическую группу, а 100, сделала бы результаты по корреляции более точными.

На момент написания работы в некоторых систематических группах было набрано менее 25 видов, поскольку в базе EMBL-банка отсутствовали геномы этих систематических групп в необходимом количестве. Отсюда дополнительная проблема в том, что сложно в принципе набрать 100 видов геномов митохондрий в некоторых редких систематических группах. Возможно в ближайшее время ситуация улучшится и станет возможным продолжить исследование корреляции GC-состава и полученных структур.

3.4 Окрашивание полученных структур по GC-составу

Следующим шагом нашей исследовательской работы было окрашивание полученных структур по GC-составу с целью выявить характерные отличия не с количеством GC-состава в систематической группе (см. пункт 3.3 результатов), но с его содержанием в конкретных участках, выраженных в точках полученных структур митохондриальных геномов.

Большая часть данной задачи заключалась в выявлении относительного количества GC-состава в точках, относящихся к подтипам трехлучевой структуры: трехлучевая с хвостом, трехлучевая с нимбом. Если говорить еще точнее, объектом повышенного интереса являются подструктуры, которые отличались: хвост и нимб соответственно.

Опираясь на результаты, представленные в окраске по прямой и обратной фазам и джанку (см. пункт 3.2 результатов рис. 8-13) далее последуют 6 соответствующих структур митохондриальных геномов, окрашенных, в этот раз, по **ГС**-составу.

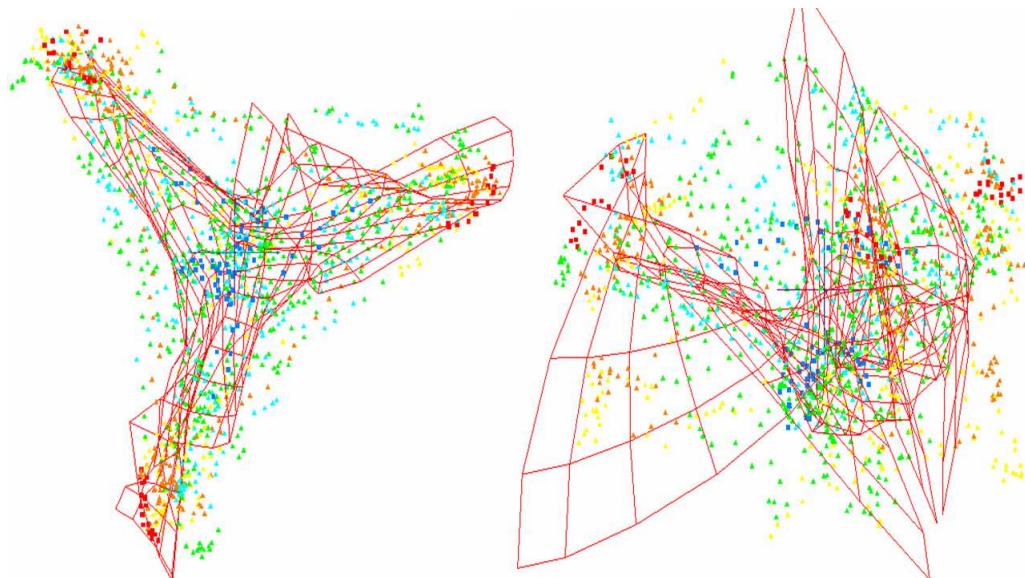


Рисунок 14 – Окрашенная по **ГС**-составу трехлучевая структура на примере *Coelodonta antiquitatis* (млекопитающие)

После проведения окраски полученных структур по **ГС**-составу, обнаруживаются следующие закономерности:

- 1) Для митохондриальных геномов наземных растений, мхов, одноклеточных водорослей, мохоподобного лишайника *Cladonia* и высших грибов характерна градиентная окраска;
- 2) Для митохондриальных геномов многоклеточных водорослей, лишайников и низших грибов характерна неявная, условная градиентная окраска;
- 3) Для митохондриальных геномов членистоногих (насекомых, паукообразных и ракообразных) характерна центральносимметричная окраска;
- 4) Для губок, моллюсков и червей (плоских, кольчатых и круглых) характерен неявный градиент;
- 5) Среди позвоночных у млекопитающих высокий **ГС**-состав в подструктуре хвост, в то время как у амфибий, рептилий и птиц в подструктуре хвоста чаще низкий **ГС**-состав. Объединяющего все систематические

группы позвоночных окраски по GC-составу обнаружено не было.

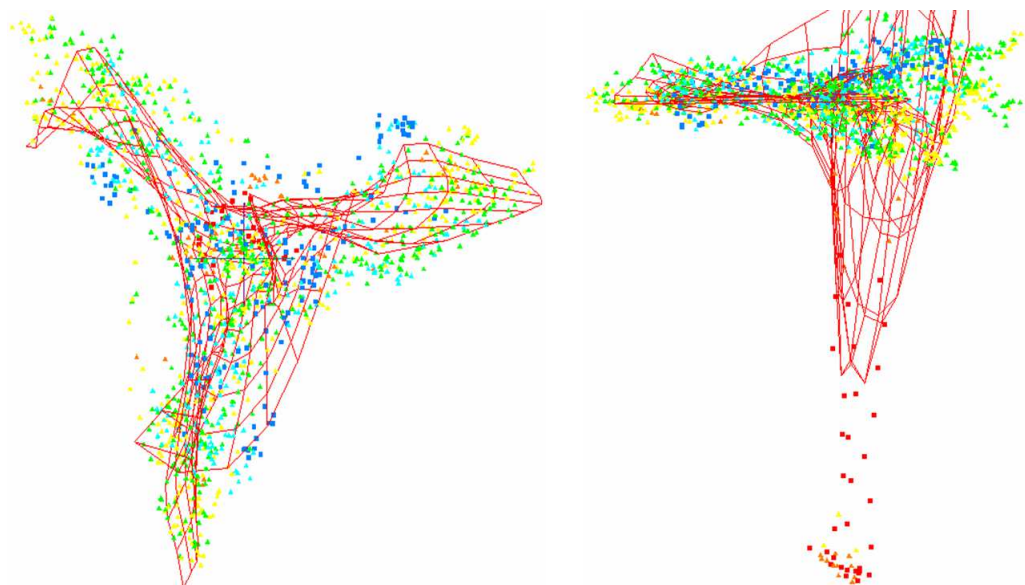


Рисунок 15 – Окрашенная по GC-составу трехлучевая структура с хвостом на примере *Cuon alpinus lepturus* (млекопитающие)

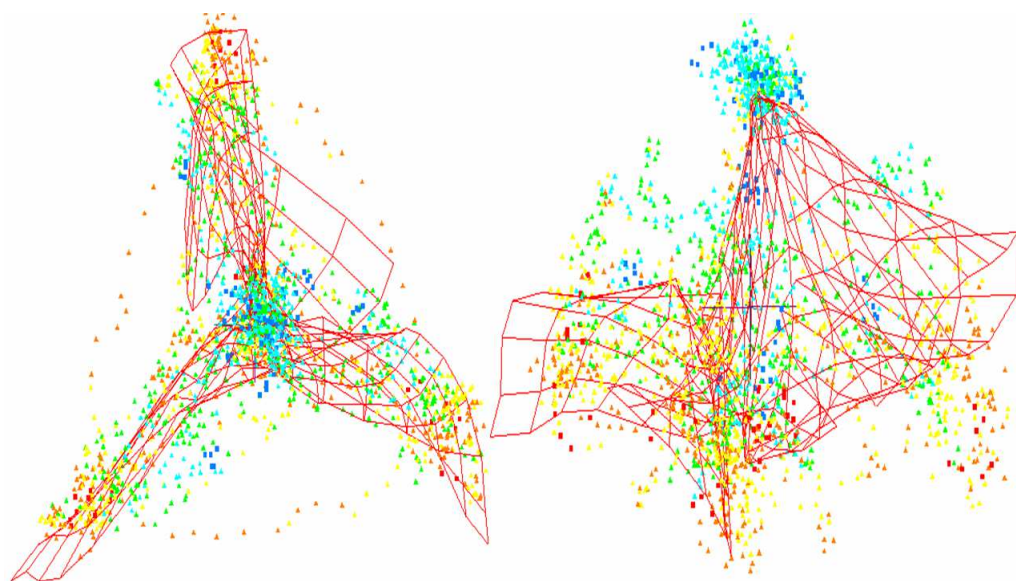


Рисунок 16 – Окрашенная по GC-составу трехлучевая структура с нимбом на примере *Chimaera fulva* (хрящевые рыбы)

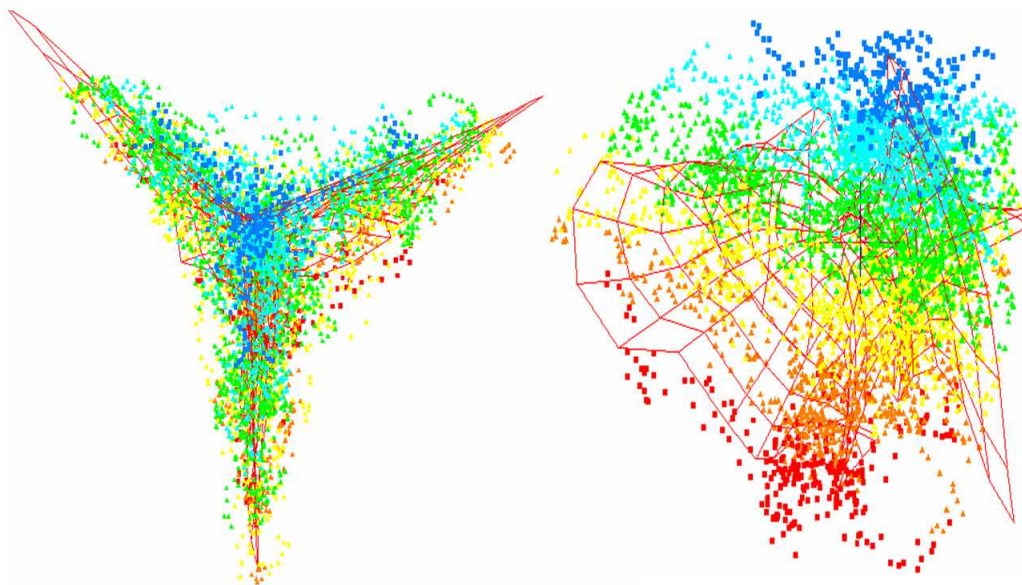


Рисунок 17 – Окрашенная по GC-составу трехлучевая структура с ядром на примере *Ganoderma lucidum* (высшие грибы)

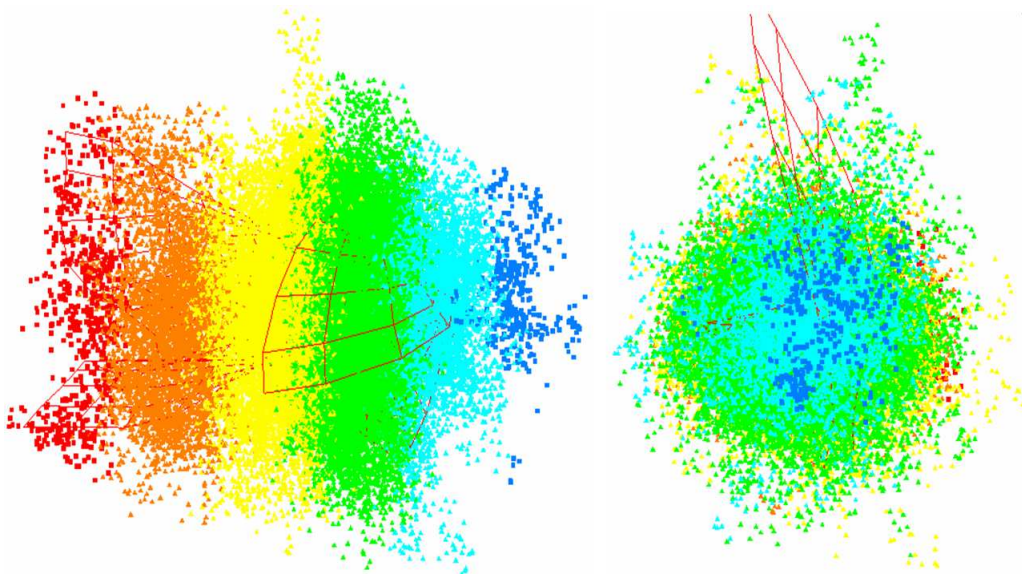


Рисунок 18 – Окрашенная по GC-составу структура клубок на примере *Oryza rufipogon* (растения)

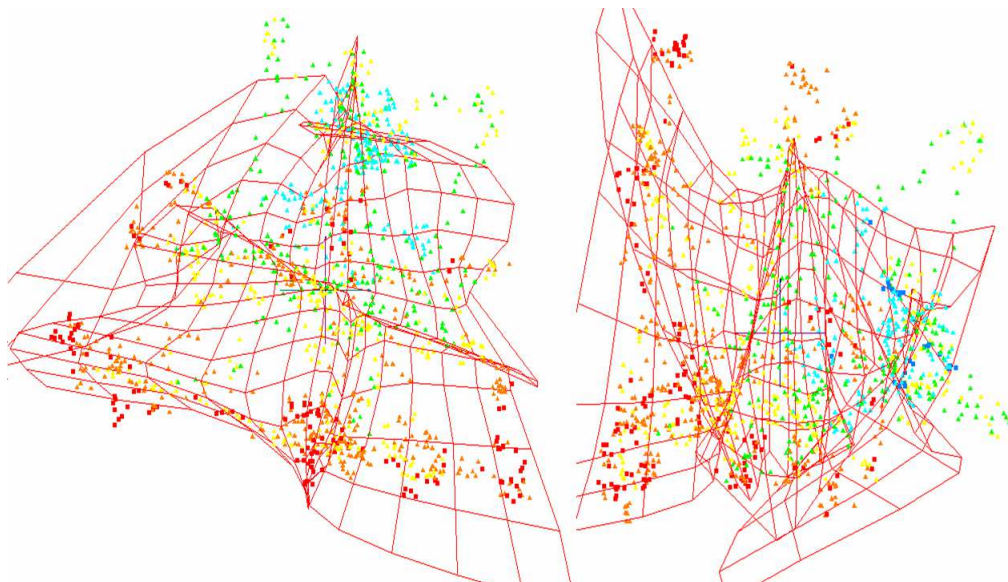


Рисунок 19 – Окрашенная по GC-составу аморфная структура на примере *Heterometrus longimanus* (паукообразные)

Помимо этого, были обнаружены иные незначительные закономерности и особенности, которые требуют дальнейшего рассмотрения в грядущих исследованиях.

ЗАКЛЮЧЕНИЕ

В ходе данной работы были выполнены все задачи, а поставленная цель была достигнута:

- 1) Была исследована закономерность между GC-составом и изучаемыми систематическими группами;
- 2) Были выявлены различные внутренние структуры митохондриальных геномов, характерные для исследуемых систематических групп;
- 3) Полученные структуры были классифицированы;
- 4) Была изучена взаимосвязь полученных структур и их отдельных компонентов от окраски по GC-составу.

Анализ результатов показал, что в большинстве своём геномы митохондрий стремятся к трёхлучевой структуре. Была обнаружена незначительная корреляция между GC-составом и систематическими группами. Так, черви обладают наименьшим GC-составом, а птицы и рептилии наибольшим. В общем и целом, были обнаружены следующие структуры: трёхлучевая структура, трёхлучевая с хвостом, трёхлучевая с нимбом, трёхлучевая с ядром, клубок и аморфная. У геномов митохондрий наблюдается огромное разнообразие выявленных структур.

С помощью окраски по GC-составу удалось обнаружить некоторые закономерности, среди которых центральносимметричная окраска (напр., у членистоногих) и градиентная окраска (напр., у растений и мхов).

Чтобы сделать результаты более точными, необходимо исследовать выборку на каждую систематическую группу больше 25, например, исследовать по 100 видов.

СПИСОК СОКРАЩЕНИЙ

1. ДНК — дезоксирибонуклеиновая кислота;
2. РНК — рибонуклеиновая кислота;
3. АТФ — аденозинтрифосфорная кислота;
4. ГТФ — гуанозинтрифосфат;
5. тРНК — транспортная РНК;
6. иРНК — информационная РНК;
7. мтДНК — митохондриальная ДНК;
8. п. н. — пары нуклеотидов;
9. мкм — микрометр;
10. GC состав (гуанин-цитозин состав) — процентный состав суммы всех гуанинов (G) и цитозинов (C) по отношению к длине исследуемого участка нуклеиновых кислот.
11. Джанк — некодирующие участки.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] *Suwanto, A.* Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4. 1 genome: presence of two unique circular chromosomes. / A. Suwanto, S. Kaplan // *Journal of Bacteriology*. — 1989. — Vol. 171, no. 11. — Pp. 5850–5859.
- [2] Молекулярная биология клетки: Учебник в 3-х томах. Т. I / Б. Альбертс, Д. Брей, Дж. Льюис et al. // «Регулярная и хаотическая динамика». — 2013. — Pp. 1318–1324.
- [3] *Wolstenholme, David R.* Animal mitochondrial DNA: structure and evolution / David R. Wolstenholme // *International review of cytology*. — Elsevier, 1992. — Vol. 141. — Pp. 173–216.
- [4] High variability of mitochondrial gene order among fungi / Gabriela Aguilera, Damien M. De Vienne, Oliver N. Ross et al. // *Genome biology and evolution*. — 2014. — Vol. 6, no. 2. — Pp. 451–465.
- [5] *Smith, David Roy.* The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? / David Roy Smith // *Briefings in functional genomics*. — 2016. — Vol. 15, no. 1. — Pp. 47–54.
- [6] *Bendich, Arnold J.* Reaching for the ring: the study of mitochondrial genome structure / Arnold J. Bendich // *Current genetics*. — 1993. — Vol. 24, no. 4. — Pp. 279–290.
- [7] *Boore, Jeffrey L.* Big trees from little genomes: mitochondrial gene order as a phylogenetic tool / Jeffrey L. Boore, Wesley M. Brown // *Current opinion in genetics & development*. — 1998. — Vol. 8, no. 6. — Pp. 668–674.
- [8] Genome structure and gene content in protist mitochondrial DNAs / Michael W. Gray, B. Franz Lang, Robert Cedergren et al. // *Nucleic acids research*. — 1998. — Vol. 26, no. 4. — Pp. 865–878.


- [9] *Taanman, Jan-Willem*. The mitochondrial genome: structure, transcription, translation and replication / Jan-Willem Taanman // *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. — 1999. — Vol. 1410, no. 2. — Pp. 103–123.
- [10] *Boore, Jeffrey L*. Animal mitochondrial genomes / Jeffrey L. Boore // *Nucleic acids research*. — 1999. — Vol. 27, no. 8. — Pp. 1767–1780.
- [11] *Burger, Gertraud*. Mitochondrial genomes: anything goes / Gertraud Burger, Michael W. Gray, B. Franz Lang // *Trends in genetics*. — 2003. — Vol. 19, no. 12. — Pp. 709–716.
- [12] Unique mitochondrial genome architecture in unicellular relatives of animals / Gertraud Burger, Lise Forget, Yun Zhu et al. // *Proceedings of the National Academy of Sciences*. — 2003. — Vol. 100, no. 3. — Pp. 892–897.
- [13] *Bullerwell, Charles E*. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants / Charles E. Bullerwell, Michael W. Gray // *Current opinion in microbiology*. — 2004. — Vol. 7, no. 5. — Pp. 528–534.
- [14] *Marande, William*. Unique mitochondrial genome structure in diplomids, the sister group of kinetoplastids / William Marande, Julius Lukeš, Gertraud Burger // *Eukaryotic cell*. — 2005. — Vol. 4, no. 6. — Pp. 1137–1146.
- [15] *Gissi, Carmela*. Evolution of the mitochondrial genome of *Metazoa* as exemplified by comparison of congeneric species / Carmela Gissi, Francesco Iannelli, Graziano Pesole // *Heredity*. — 2008. — Vol. 101, no. 4. — Pp. 301–320.
- [16] Divergence of the mitochondrial genome structure in the apicomplexan parasites, *Babesia* and *Theileria* / Kenji Hikosaka, Yoh-ichi Watanabe, Naotoshi Tsuji et al. // *Molecular biology and evolution*. — 2010. — Vol. 27, no. 5. — Pp. 1107–1116.

- [17] *Gray, Michael W.* Mitochondrial evolution / Michael W. Gray, Gertraud Burger, B. Franz Lang // *Science*. — 1999. — Vol. 283, no. 5407. — Pp. 1476–1481.
- [18] *Hikosaka, Kenji.* Diversity of mitochondrial genome structure in the phylum *Apicomplexa* / Kenji Hikosaka, Kiyoshi Kita, Kazuyuki Tanabe // *Molecular and biochemical parasitology*. — 2013. — Vol. 188, no. 1. — Pp. 26–33.
- [19] *Cameron, Stephen L.* Insect mitochondrial genomics: implications for evolution and phylogeny / Stephen L. Cameron // *Annual review of entomology*. — 2014. — Vol. 59. — Pp. 95–117.
- [20] *Садовский, М. Г.* Восьмикластерная структура геномов хлоропластов наземных растений / М. Г. Садовский, М. Ю. Сенашова, А. В. Малышев // *Журнал общей биологии*. — 2018. — Vol. 79, no. 2. — Pp. 124–134.
- [21] *Gorban, Alexander N.* Seven clusters in genomic triplet distributions / Alexander N. Gorban, Andrei Y. Zinovyev, Tatyana G. Popova // *In silico biology*. — 2003. — Vol. 3, no. 4. — Pp. 471–482.
- [22] *Gorban, Alexander N.* Four basic symmetry types in the universal 7-cluster structure of microbial genomic sequences / Alexander N. Gorban, Tatyana G. Popova, Andrei Y. Zinovyev // *In silico biology*. — 2005. — Vol. 5, no. 3. — Pp. 265–282.
- [23] *Gorban, Alexander.* Codon usage trajectories and 7-cluster structure of 143 complete bacterial genomic sequences / Alexander Gorban, Tatyana Popova, Andrey Zinovyev // *Physica A: Statistical Mechanics and its Applications*. — 2005. — Vol. 353. — Pp. 365–387.
- [24] The EMBL-EBI bioinformatics web and programmatic tools framework / Weizhong Li, Andrew Cowley, Mahmut Uludag et al. // *Nucleic acids research*. — 2015. — Vol. 43, no. W1. — Pp. W580–W584.

- [25] *Косарев, Р. Е.* Пространственная структура митохондриальных геномов растений и животных / Р. Е. Косарев, М. Ю. Сенашова, М. Г. Садовский // *Нейроинформатика, её приложения и анализ данных.* — 2019. — Рр. 58–66.

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт фундаментальной биологии и биотехнологии
Кафедра биофизики


УТВЕРЖДАЮ:
заведующий кафедрой

 В. А. Кратасюк
«22» июня 2020 г.

БАКАЛАВРСКАЯ РАБОТА

06.03.01 Биология

ИЗУЧЕНИЕ ВНУТРЕННЕЙ СТРУКТУРИРОВАННОСТИ ГЕНОМОВ МИТОХОНДРИЙ НЕКОТОРЫХ СИСТЕМАТИЧЕСКИХ ГРУПП

Руководитель:  22.06.2020 д.ф.-м.н., проф. М. Г. Садовский
дата, подпись уч. степень, должность

Выпускник: 22.06.20  Р. Е. Косарев
дата, подпись

Красноярск 2020